

An Examination of Different Methodological Approaches in Student School Behavior Research: The Issue of the Incomparability of Student Self-assessments¹

Stanislav Bendl, Hana Vonkova, Ondrej Papajoanu,
Eva Vankatova

Charles University, Faculty of Education, Department of Education

Received 4th July 2018 / final version received 30th October 2018 /
/ accepted 1st November 2018

Abstract: The issue of school educational outcomes measurement is of great concern to both researchers and practitioners. We can distinguish two main types of outcomes: outcomes in the domain of academic achievement (e.g., mathematics, information and communication technologies, and history) and outcomes in the behavioral domain (school discipline). Both types of outcomes are assessed and graded in schools. However, if we were to let different teachers assess the same students' knowledge and skills, their assessment would frequently differ and the same applies to their assessments of student behavior. The following question arises: How accurately do we measure school educational outcomes? In our study we aim to address the following issues: (a) describe the most commonly used methodological approaches to student school behavior measurement and discuss their strengths and weaknesses. Specifically, we focus on school documentation analysis, interviews, observations, and questionnaire surveys. The section about school documentation analysis focuses on empirical analysis of the selected school documents of particular Czech schools; (b) propose an innovative approach to student school behavior measurement combining student self-reports and peer-reports with the anchoring vignette method to enhance data comparability.

Keywords: school discipline, academic achievement, self-assessment, bias, anchoring vignette method

This study deals with the issue of student school behavior measurement. In general, the issue of school educational outcomes measurement is a key issue in the literature. We can distinguish two main types of outcomes: outcomes

¹ This paper was enabled by the project *Factors influencing the ICT skill self-assessments of upper-secondary school students* (17-02993S) funded by the Czech Science Foundation.

in the domain of academic achievement (e.g., mathematics, information and communication technologies, and history) and outcomes in the behavioral domain. Both types of outcomes (academic achievement and behavior) are assessed and graded in schools. However, if we were to let different teachers assess the same students' knowledge and skills, their assessment would frequently differ. For example, Bendl (1987) let different teachers assess the same students' work which was indicative of students' performance in Czech language classes (e.g. different types of grammar exercises, essays) and found that some of the participating teachers assessed the same students' work differently. There was a difference of up to two points on a Czech five-point grading scale (1 = *excellent*, 2 = *very good*, 3 = *good*, 4 = *sufficient*, 5 = *insufficient*). A similar situation occurs in teachers' assessments of student discipline.

The following question arises: How accurately do we measure school educational outcomes?

Here we focus on the methods/approaches suggested for student school discipline measurement and examine their specific properties. Special attention is paid to students as an information source. Even though students in schools are typically the "objects" of assessment, they are active participants in the educational process and can provide a valuable perspective on a variety of educational outcomes. Student self-assessments (typically questionnaires with rating scales) are frequently employed in educational research and have the potential to contribute to the measurement of both the academic and behavioral outcomes of the educational process. The combination of student self-assessment and peer-assessment appears to be a promising approach: a student is not only assessing his/her behavior, but also the behavior of his/her peers (classmates). However, it has been recognized that both student self-assessments and peer-assessments can be biased by the differences in scale usage between different respondents. We also address this issue and suggest its potential solution.

Our study has the following specific aims:

- 1) Review the most commonly used *methodological approaches* (school documentation analysis, interviews, observations, questionnaire surveys) to student school behavior measurement and discuss their strengths and weaknesses and the conditions under which they can be used.

In particular, we:

- discuss the issue of using *school documentation* to measure the prevalence of school (mis)behavior in the context of U.S. and Czech schools. The current use of office discipline referrals (ODRs) in school behavior research is described for the U.S. With regards to Czech schools, we discuss the use of school behavior grading and official sanctioning as indicators of school misbehavior levels. In the same sub-section we also provide an empirical analysis of the selected school documents of particular Czech schools. Both the strengths and weaknesses of these indicators are pointed out;
 - describe several examples of studies of student school behavior where *interviews* were employed. The role of interviews in providing in-depth information on school misbehavior incidents and in the examination of school behavioral interventions is highlighted. The strengths and weaknesses of interviews in large-scale research are described;
 - provide a brief overview of the two basic *types of observation*: naturalistic observation and systematic direct approaches. Several ways of data recording (A-B-C recording, event recording, time-sampling interval recording etc.) are introduced together with their main characteristics. The well-established use of observations in student school behavior research is illustrated. However, some major weaknesses to using observations in large-scale research whose goal is determining misbehavior prevalence are emphasized;
 - discuss the strengths of using *questionnaires* in large-scale surveys and illustrate their wide-spread use in research into student school discipline. We examine the issue of the low level of agreement between different informants on student school behavior (parents, teachers, peers, and students themselves) and provide an overview of the potential strengths and weaknesses related to the use of different informants. We also draw attention to the problems with the (in)comparability of data obtained using questionnaires with ratings scales due to respondents' differential scale usage.
- 2) Propose an *innovative approach* to student school behavior measurement combining student self-reports and peer-reports while adjusting for differences in scale usage among respondents.

1 Measuring approaches

In this section we describe the most commonly used methodological approaches in student school behavior research: school documentation analysis, interviews, observations, and questionnaires. For each we provide example studies illustrating their use in school discipline research. In the case of school documentation, two specific areas are focused on in detail: (a) a sub-section dealing with the use of office discipline referrals in the context of U.S. schools, (b) a sub-section dealing with the use of school behavior grading and sanctioning in the context of Czech schools. The latter sub-section also contains an empirical analysis of the documentation of selected Czech schools which supports the line of argument being provided there. The major strengths and weaknesses of each of these approaches are pointed out. The U.S. system was chosen because of the vast amount of literature regarding student school discipline that is published in the context of U.S. schools and the frequent use of ODRs as a school behavior indicator. The Czech system was chosen because it is both familiar and relevant to the intended readers.

1.1 School documentation

One of the approaches that is frequently used in studies of school discipline when identifying the level of disciplinary problems in schools, is the analysis of school documentation.

Office discipline referrals – strengths and weaknesses (in the context of U.S. schools)

Office discipline referrals (ODR) are frequently employed as an indicator of the overall disciplinary climate, particularly in studies conducted in the U.S. ODR can be defined as a situation when: (a) a student violated some of the school rules or norms, (b) his/her problem behavior was noticed by some member of the school staff, (c) the event resulted in a consequence delivered by administrative staff who produced a written record describing the whole event (Sugai et al., 2000). ODR forms² usually contain such

² It is possible to find examples of ODR forms online (Todd & Horner, 2006). Sometimes, an ODR form for minor infractions and an ODR form for major infractions are distinguished. It is also possible to find ODR forms containing further information concerning the incident like the information about other people involved in the incident or the possible motivation for the behavior.

information as: time, date, name of the student who violated the rules, name of the referring teacher, location of the incident, type of problem behavior, or type of consequence that was delivered to the student. It has been suggested that ODR data are a valuable source of information for schools concerning their students' school behavior and can be used for data-based decision making in terms of school prevention efforts (Irvin et al., 2006; Sugai et al., 2000). There are also sophisticated computer applications such as *School Wide Information System* (SWIS) for entering, organizing, managing, and reporting the ODR data to be used in school decision making (Irvin et al., 2006). According to Irvin et al., ODR data in SWIS can be used not only for internal school decision-making concerning school discipline, but also to plan the support provided to individual students, to report discipline data to the district/state, and to aggregate and interpret data across different schools. Standardized SWIS reports summarize the following information: (a) ODR per day per month for the whole school, (b) ODR per type of problem behavior, (c) ODR per student, (d) ODR per location in the school, (e) ODR per time of day (Irvin et al., 2006). ODR data is also frequently used as an outcome measure in studies examining the impact of behavioral interventions in schools (e.g. Bohannon et al., 2006; Luiselli et al., 2005; McCurdy, Mannella, & Eldridge, 2003).

The major advantage of using ODR data as an indicator of school discipline levels is that they are already collected in many schools (Sugai et al., 2000) so they can serve as an efficient source of information for the school itself. The data are collected on a regular basis allowing the identification of long-term trends in school discipline levels. Also, the use of computer applications to record ODR (or similar records of discipline infractions) could be utilized by the researchers – the readily available data from different schools may be collected in a central database and then analyzed for research purposes. However, there are some limitations to using ODR data as a school discipline indicator. First, each school defines and applies referral procedures in a unique manner, that is the same student behavior may be reacted to differently by teachers in different schools (Sugai et al., 2000). ODR can also be administered differently by teachers in the same school depending on their tolerance level and their skills at handling student behavior (Morrison & Skiba, 2001). ODR data might also be biased by other factors – it would appear plausible that the probability of a student receiving an official sanction like an ODR might be influenced by the relationship

between the teacher and a particular student. Also, official sanctions like an ODR capture only those incidents that reach a certain level of severity thus not providing information about the less severe (but maybe very prevalent) types of student misbehavior. Lastly, for the ODR or any official sanction to be administered, the school staff has to notice the behavior in the first place. However, since some of the types of student misbehavior are meant to remain hidden to teachers (e.g. students cheating on exams), it is improbable that the number of disciplinary sanctions would correspond to the actual prevalence of the behavior. A summary of the strengths and weaknesses of office discipline referrals is given in Table 1.

Table 1

Office Discipline Referrals – A Summary of the Strengths and Weaknesses

Strengths	Weaknesses
Easy use of already collected data (already collected in many schools, use of electronic data systems)	The definitions and applications of referral procedures differ across schools
Collected on a regular basis (allowing the examination of trends in behavior)	Differences in ODR administration based on teachers' skills and tolerance levels
Many types of information about disciplinary incidents (time, place, type of misbehavior etc.)	Does not capture less severe rule violations
	School staff might not notice some rule violations

School behavior Grading and Sanctioning – strengths and weaknesses (in the context of Czech schools)

In the Czech Republic, a similar kind of school documentation data could be used to measure school misbehavior levels in schools. Czech schools can formally sanction the students for their misbehavior using three different types of official reprimands that are graded by their severity: (a) an official reprimand of lesser severity administered by a classroom teacher (*napomenutí třídního učitele*), (b) an official reprimand of medium severity administered by a classroom teacher (*důtka třídního učitele*), (c) an official reprimand of higher severity administered by a school principal (*důtka ředitele školy*). The information about the administration of these official sanctions is recorded in a school's documentation. We could explore the data about the number of these formal sanctions that were administered by particular schools to determine the overall school misbehavior level. According to the Czech School

Inspectorate³ (CSI), in the school year 2015/2016 77.3% of Czech basic schools⁴ (primary level) and 95.8% (lower secondary level) administered an official reprimand of lesser severity to a student, 66.8% (primary level) and 95.4% (lower secondary level) an official reprimand of medium severity, and 45.4% (primary level) and 89.8% (lower secondary level) an official reprimand of higher severity (CSI, 2017).

Also, students in Czech schools receive a formal report of their school achievement in different subjects, semi-annually. Their performance in each subject is summarized and represented by a single final grade. Part of this final report is also a final grade in the domain of school behavior (discipline). Unlike achievement in particular subjects, which is graded on a five-point scale (1 = *excellent*, 2 = *very good*, 3 = *good*, 4 = *sufficient*, 5 = *insufficient*), student school behavior is graded on a three-point scale (1 = *very good* [best grade], 2 = *satisfactory*, 3 = *unsatisfactory* [worst grade]). It is also possible to analyze student final grades in the domain of school behavior to determine the overall discipline level at particular schools. In the school year 2015/2016, 23.1% of Czech basic schools (primary level) and 76.3% (lower secondary level) graded a student with grade 2 (satisfactory) in the domain of school behavior and 5.9% (primary level) and 43.7% (lower secondary level) graded a student with grade 3 (unsatisfactory) in the domain of school behavior (CSI, 2017).

The advantage of school documentation as a data source is that the data is readily available since schools record both student grades in the domain of behavior and the official disciplinary sanctions that are administered to students. Nowadays many schools use electronic systems to record student grades, attendance, and other information, involving their final grades and administered disciplinary sanctions. A wide-spread example of this electronic system in the Czech Republic is *Bakaláři*⁵. Having the data accessible in electronic form can further simplify the use of the data and their analysis. Indeed, the system allows schools to do some basic analysis of school

³ The Czech School Inspectorate is an administrative body of the Czech Republic and an organizational component of the state. Web pages: <http://www.csicr.cz>

⁴ The sample consisted of 3 464 Czech basic schools (primary and lower secondary level). The percentages represent schools that administered at least one reprimand of a particular type to a student during the school year 2015/2016. The same applies to student grades in the domain of school behavior discussed later in this section. For more information see CSI (2017).

⁵ <https://www.bakalari.cz/>

behavior data (e.g. basic summaries, graphs). It must be noted, however, that even though schools record the data on student behavior, it is not collected in any single central database that would allow large-scale analysis.

However, we should note that some statistical information about student school behavior is available. The former Institute for Information on Education (IIE) conducted so-called *Quick Surveys*⁶ where a representative sample of 4000 schools⁷ (resp. their principals) were surveyed on various topics, some of which were related to student school behavior (bullying, aggression). The principals were, for example, asked about the number of times students had come to school with different types of weapons or about the frequency of bullying incidents during that particular school year (IIE, 2007, 2008).

Nowadays, the Czech School Inspectorate surveys schools about different topics through the *InspIS* electronic system, including topics related to student school behavior. The school principals might use the information from the *Bakaláři* system to fill in the questionnaires of the CSI. Also, the CSI visits a number of schools each year to conduct in-depth school inspections. The scope of information on student school behavior collected by the CSI differs from year to year, but every year at least some basic indicators are collected. In the CSI annual and thematic reports, it is possible to find some summarizing data about student school behavior. For example, in the annual reports for the school years 2015/2016 and 2016/2017⁸ (CSI 2016, 2017) it is possible to find the percentages of schools that had to deal with diverse types of risky behavior (truancy, bullying, vandalism etc.). In the school year 2015/2016, 41% of schools reported dealing with bullying in the previous school year, while in the school year 2016/2017, 35.3% of schools reported dealing with bullying in the previous school year. Still, the percentages must be interpreted with caution in terms of the school misbehavior level measurement, since they do not reflect the number or the severity of the

⁶ In Czech, these are called *Rychlá šetření* conducted by Ústav pro informace ve vzdělávání.

⁷ For more information on the sample see IIE (2007).

⁸ The data on school behavior published in the reports are based on the samples of 787 basic schools (2015/2016) and 867 basic schools (2016/2017). For more information about the composition of the sample, see the respective annual reports (CSI, 2016, 2017). During school inspections at these schools, the inspectors investigated whether the school dealt with at least one incident of a range of different types of risky behavior during the previous school year. So, the data in the annual report for school year 2015/2016 actually correspond to the year 2014/2015 and the same applies for the 2016/2017 report.

incidents. For example, at one school there might have been a single low severity level incident of bullying during the school year. At another school, there might have been a number of high-severity level bullying incidents during the school year. Both schools, however, would be included into the statistics as schools where bullying took place without any differentiation made between them.

As with the previously mentioned ODRs, there are several severe limitations to using both student grades in the domain of discipline and the formal sanction data for large-scale school discipline measurement and, for example, the comparison of different (types of) schools. Every Czech school is mandated to have an internal document⁹ regulating its functioning in different domains (e.g. student rights and responsibilities, expected norms of behavior, student safety, or standards for student assessment). It also covers the issue of school disciplinary sanctions and the specifics of their administration. However, the actual content of this document varies across schools which apply different approaches to dealing with student misbehavior and its grading/sanctioning.

The school codes of particular schools specify how student behavior is graded. However, the exact specifications of the declared standards for student behavior grading differ across schools. Many schools include only very general descriptions (see Table 2) of student behavior and their correspondence to a particular grade, into their codes, e.g. Základní škola a mateřská škola Bílá / Basic school¹⁰ and Kindergarden Bílá (ZŠ a MŠ Bílá). These general descriptions as such do not provide very detailed information about the concrete standards which particular schools apply when assessing student behavior. It might be the case that two schools differ in their actual assessment of student behavior, even though they both formally adhere to these very general descriptions. Also, the school codes often state that when grading student behavior, student age, moral and cognitive development should be considered (ZŠ a MŠ Bílá, 2017). This introduces further “inaccuracy” into student behavior grading (in terms of objective school misbehavior level measurement) since, as the above-mentioned statement suggests, the same behavior of two students might be judged differently based on their developmental level.

⁹ In Czech this document is called *školní řád* (translated as the school code; Průcha, 2005). It is a set of rules and regulations governing the functioning of a school.

¹⁰ In Czech school system basic school typically covers primary and lower secondary level (age 6 to 10 and 11 to 15 respectively).

Some schools provide more concrete descriptions of the standards for student behavior grading, specifying some of the violations that correspond to a particular grade (e.g. The Bělský Les Kindergarten and Basic School, Ostrava / Základní škola a mateřská škola Ostrava – Bělský Les, 2017). However, when the schools specify their grading standards in more detail, the differences in grading standards among schools become more obvious. A good example is the number of unexcused absences for which a student is given a particular grade in the domain of school behavior. In the Jan Werich Basic School / Základní škola Jana Wericha (2011, 2017), 3 to 10 unexcused absences correspond to grade 2 in the domain of behavior. However, in the Želenice Basic School / Základní škola Želenice (2012), 11 up to 20 unexcused absences (lessons) are assessed as grade 2.

Concerning school disciplinary sanctions, the situation is very similar. The school codes of particular schools specify how these sanctions are administered. However, as with student grading in the domain of behavior, these specifications differ greatly across schools (see Table 3). For example, the Sázavská Basic School / Základní škola Sázavská (2013) does not specify the standards for the administration (i.e. the types of misbehavior for which a particular sanction would be administered) of these sanctions at all. Some schools provide a basic description of the misbehaviors for which a formal sanction of a particular degree of severity will be administered. For example, the Petřiny – North Basic School / Základní škola Petřiny – sever (2015) administers the least severity level sanction for minor misbehaviors, forgetting school equipment, and classroom disruption. The medium severity level sanction is administered for forgetting homework or a student's report book, repeated late arrivals, bad working morale, inappropriate behavior of a lesser degree.

These descriptions, even though they indeed specify the behavior for which a certain sanction can be administered, are very general and it might be difficult to say where exactly is the borderline between “minor misbehaviors” and “inappropriate behavior of a lesser degree”. It is also stated that each rule violation is judged individually, all the circumstances are taken into consideration, and also the consequences of the sanction administration are considered. The vague definition of student misbehavior and the involvement of such a broad scope of circumstances into the sanction administration cast some doubt about their “accuracy” in terms of student

misbehavior measurement. If the categories of student misbehavior are only loosely defined, different teachers can administer different sanctions for the same behavior because each one will judge it differently. Also, the presence/absence of some extenuating or other contextual circumstances can result in administering different sanctions for the equivalent incidents of rule violation.

Table 2

Examples of Student School Behavior Grading Standards Defined in the School Codes of Selected Czech Schools

School	ZŠ a MŠ Bílá	ZŠ a MŠ Ostrava – Bělský Les
How behavior grading standards are set	A general definition of student behavior	A definition of student behavior including a complex list of particular examples
A description of the rule violations corresponding to grade 2 (the lower grade for school behavior)	Grade 2 (satisfactory) The behavior of a student is not in line with the norms of behavior defined by the school code. The student committed a serious violation against the norms of proper behavior or the school code or he/she repeatedly commits less serious violations. Usually he/she commits further rule violations even after he/she has been sanctioned by an official reprimand of medium severity administered by a classroom teacher and disrupts school educational efforts. He/she puts his/her own health or the health of others at risk.	Grade 2 (satisfactory) The behavior of a student is not in line with the norms of behavior defined by the school code. The student commits serious rule violations or repeatedly commits less serious violations, while not being susceptible to educational efforts. Grade 2 in the domain of school behavior corresponds to, for example, 2 days of unexcused absence (or repeated absence of up to 12 lessons or 1 day of unexcused absence + other violations), hurting peers, bullying, xenophobic behavior, leaving the school building without permission during instruction time, theft, rude and vulgar behavior towards peers and school staff (see School Law, § 31, par. 3), repeated lying, deceit, vandalism, consummation of alcohol and smoking cigarettes on the school premises or during school-organized events.

Note. We provide the translations of the description for grade 2. In the Czech Republic, student school behavior is graded on a three-point scale (1 = *very good* [best grade], 2 = *satisfactory*, 3 = *unsatisfactory* [worst grade]). For the description of other grades see the respective school codes.

Some schools developed quite sophisticated point systems where students are given negative points for inappropriate behavior. For a certain number of these points, the official sanctions of different severity are administered. For example ZŠ a MŠ Bílá (2017) assesses such rule violations as late arrival, forgetting a student's report book, or using electronic devices such as mobile phones or tablets in school with 1 point. An official reprimand of lesser severity administered by a classroom teacher will be administered for 6 points. Another example is ZŠ Jana Wericha (2017) which defines over 20 types of misbehavior and states the exact number of points or a point interval for committing each of these types of misbehavior, ranging from late arrival to bullying. An official reprimand of lesser severity administered by a classroom teacher will be administered for 4 points. These more clearly defined descriptions of a school's official sanction administration policies provide a better idea of how these schools administer these sanctions and what are the standards for student behavior.

However, as the standards for sanction administration become more explicit, the differences between particular schools become more evident. For example, at ZŠ a MŠ Bílá, the use of mobile phones, tablets, or other electronic devices in school is sanctioned by 1 point. At ZŠ Jana Wericha, the use of a mobile phone during lessons is sanctioned by 2 points. Thus, hypothetically, at ZŠ a MŠ Bílá a student can use a mobile phone during a lesson six times before he/she receives an official reprimand of lesser severity administered by a classroom teacher (6 points), while at ZŠ Jana Wericha a student can use a mobile phone during a lesson only twice before he/she is administered the very same sanction (4 points).

Table 3
Examples of Disciplinary Offences Sanctioning Defined in the School Codes of Selected Czech Schools

School	ZŠ Sázavská	ZŠ Peříný – sever	ZŠ a MŠ Bílá	ZŠ Jana Wericha
Point system	Non-existing	Non-existing	Existing (basic)	Existing (complex)
Disciplinary offences corresponding to sanctions	No definition	General definition	Definition elaborated by a point system	Definition elaborated by a complex point system
Sanction specification	Sanctions are only listed	Sanctions are specified only generally	Sanctions are specified mainly by the number of points	Sanctions are specified by the number of points
Misbehavior sanctioning as defined in school codes	(a) An official reprimand of lesser severity administered by a classroom teacher;	(a) An official reprimand of lesser severity administered by a classroom teacher – minor misbehaviors, forgetting school equipment, and classroom disruption;	(a) An official reprimand of lesser severity administered by a classroom teacher – 6 points.	(a) An official reprimand of lesser severity administered by a classroom teacher – 4 points.
	(b) An official reprimand of medium severity administered by a classroom teacher;	(b) An official reprimand of medium severity administered by a classroom teacher – forgetting homework or a student's report book, repeated late arrivals, bad working morale, inappropriate behavior of a lesser degree;	(b) An official reprimand of medium severity administered by a classroom teacher – 12 points. (c) An official reprimand of higher severity administered by a school principal – 18 points.	(b) An official reprimand of medium severity administered by a classroom teacher – 8 points.

Table 3 (continued)

School	ZŠ Sázavská	ZŠ Petřiny – sever	ZŠ a MŠ Blá	ZŠ Jana Wericha
	(c) An official reprimand of higher severity administered by a school principal.	(c) An official reprimand of higher severity administered by a school principal – a higher rate of the previously mentioned violations, aggressive behavior, bullying, hurting younger and weaker peers, vandalism, low levels of truancy.	For each of the following rule violations, a student will earn 1 point. – not wearing internal footwear inside the school building; – forgetting a student's report book or note book; – using a mobile phone, tablets, or other electronic devices in school. <i>(There are a few other examples, please see the school code.)</i>	(c) An official reprimand of higher severity administered by a school principal – 12 points. Violations: – late arrival (repeatedly): 1 point; – not having school equipment (2x): 1 point; – not doing homework (2x): 1 point; – using a mobile phone during lessons: 2 points; – throwing objects out of windows: 2 points; – lying and cheating: 4 to 12 points; – bullying (any form): 18 points; – theft: 4 to 18 points. <i>(There are many other examples, please see the school code.)</i>

Note. In the table there are sometimes only excerpts from the school codes or shortened forms of their passages that illustrate the differences between schools since the complete school codes would be too long to present in the paper. An interested reader can see the original school codes for more complex information.

To sum up, as with the student grades in the domain of school behavior, using disciplinary sanctions data for the large-scale measurement of school misbehavior levels has some severe limitations. Some schools define their sanction administration policy only very loosely (if at all) providing space for variability between teachers in sanction administration. Also, the circumstances of particular incidents are taken into consideration when administering the official sanctions, making these sanctions more prone to inaccuracy as the indicators of school misbehavior level. Some schools define their sanction administration policy more clearly and use for example well-structured point systems. However, the comparison of these point systems indicates significant differences between schools in terms of their standards for official sanction administration. Lately, the media has also informed us that some teachers consider these official sanctions ineffective and, therefore, they do not use them at all (iDnes.cz, 2018). This further supports the notion that there might be notable differences in the administration of these sanctions not only between schools, but also between individual teachers, who might be inclined to use these official sanctions to a different degree. In other words, two teachers might give different sanctions for the same student misbehavior: one might administer an official sanction while the other could use other ways of disciplining the student. Also, as was already mentioned with ODRs, some forms of student misbehavior are intended to remain hidden to teachers (bullying, cheating) so the number of administered disciplinary sanctions might not be representative of the real prevalence of these types of behavior. For a summary of the strengths and weaknesses of school behavior grading and sanctioning see Table 4.

Table 4

School Behavior Grading and Sanctioning – A Summary of the Strengths and Weaknesses

Strengths	Weaknesses
Disciplinary indicators are already recorded by schools (so they can easily be used for research purposes)	Vague definitions of student behavior grading and sanctioning procedures Different student behavior grading and sanctioning procedures across schools Does not capture less severe rule violations School staff might not notice a rule violation

1.2 Interviews

Interviews can provide in-depth information about various forms of student misbehavior. Researchers into school discipline can employ interviews especially when they focus on the unique perspectives of the participants, detailed accounts of disciplinary incidents, or the emotional/behavioral responses to these incidents. For example, Casey-Cannon, Hayward, and Gowen (2001) used interviews to examine middle-school girls' experiences of peer victimization. Based on the interviews, they were able to document several detailed accounts of peer victimization that the girls experienced during their studies. The girls described their emotional reactions to the incidents, providing an insight into their feelings related to the victimization experience. The behavioral responses of the girls to the victimization were examined together with the information about how they perceived the appropriateness of their response and how they would respond if victimization reoccurred. Interviews also allowed the researchers to examine the impacts of the victimization experience on the girls' self-image and peer relationships, i.e. how they felt about themselves and how it impacted their friendships and acquaintanceships. As pointed out by Crothers and Levinson (2004) in their bullying assessment review, the advantage of using interviews is also that school children have an opportunity to speak about issues regarding bullying that may not be typically addressed in other formal assessment measures.

Interviews can also be used to examine the perceptions of school staff regarding school behavioral interventions (e.g. Lindsey, 2008). Lindsey acknowledges that in the process of the diffusion of innovations (e.g. new behavioral interventions in schools) the true quality of an innovation is not as important as the user's perception of its worth. She conducted interviews with teachers, principals and other relevant personnel to find out what characteristics of *Positive Behavior Interventions and Supports* (PBIS) – an educational innovation that promotes socially appropriate behaviors among students – affect their adoption by schools. Some of the important aspects influencing the diffusion process are: (a) relative advantage – the extent to which an innovation is viewed as better than what is currently being used; (b) compatibility – the degree to which others perceive the innovation to be congruent with the current norms, values, beliefs, or experiences; (c) complexity – the degree of sophistication associated with the innovation (i.e. innovations too complicated to understand and operate will be adopted

at a slower rate); (d) trial-ability – how easily an innovation can be piloted on a small scale to determine its benefits; (e) observe-ability – how obvious the advantages of an innovation are to potential adopters. Thus, examining the perceptions of the “adopters” of behavioral intervention innovations through interviews seems very helpful for both researchers and practitioners. Nastasi and Schensul (2005) strongly emphasize the role of qualitative research (where an in-depth interview is one of the primary methods) in school intervention research especially when it comes to documenting challenges in intervention implementation, examining cultural or contextual factors influencing intervention effectiveness, or the social or ecological validity of interventions.

Table 5

Interviews – A Summary of the Strengths and Weaknesses

Strengths	Weaknesses
In-depth information about various forms of student misbehavior	High time/personnel costs
The unique perspectives of the participants	Low practicability for large-scale assessment
Detailed accounts of disciplinary incidents	A high risk of bias decreases the comparability of data (e.g. different responses from participants elicited by different interviewers)
The emotional/behavioral responses of students to these incidents	
Flexibility (interviews allow to ask further supplementary questions based on a respondent's previous answers)	

Even though interviews are a valid method in school discipline research, there appear to be several limitations to using them to determine the actual level of student misbehavior in schools. Crothers and Levinson (2004) state as the foremost weakness of interviewing the time investment necessary to meet with students in order to adequately sample the entire student population. Also, they mention that different interviewers may elicit a variety of responses from children and that there is a significant danger of bias caused by the preconceptions or viewpoints of the interviewers. Both the low practicability of interviewing for a large-scale assessment and the high risk of biases decreasing the comparability of the data, limit the use of interviewing in the measurement of student school misbehavior levels. However, interviews can be used as a preliminary step in the construction of questionnaires on student school misbehavior (Ding et al., 2008, 2010). The use of interviews for small-scale in-depth studies or as a qualitative

“supplement” to quantitative data certainly has its place in school discipline research. For a summary of the strengths and weaknesses of interviews see Table 5.

1.3 Observations

Observations of student behavior are a great source of a large amount of data on student behavior. Hintze, Volpe, and Shapiro (2002) place it among the most widely used assessment procedures of school psychologists.

Observation of student behavior is also well-established as a tool for research into school behavioral interventions and their effectiveness. For example, systematic observations are often conducted to determine the differences in the occurrence of the targeted observable behavior before and after the intervention has been implemented (e.g. Barrish, Saunders, & Wolf, 1969; Campbell & Anderson, 2011). Some typical examples of studies using observations of student behavior are those made into *Good Behavior Game* (a universal classroom behavioral intervention, Barrish et al., 1969) which examine the intervention’s effectiveness in reducing student disruptive behavior (e.g. Flower et al., 2014; Lannie & McCurdy, 2007; Mitchell et al., 2015).

Hintze et al. (2002) provide an overview of the best practices for observations of student behavior. Here we provide a shortened overview of these types of observations together with their main characteristics.

In the case of *naturalistic observation*, the observer records behavioral events in their natural setting (e.g. a classroom) and observes all that is going on there, without any specific behavior in mind. The most common way of recording the events is keeping anecdotal or descriptive records of the behaviors that appear important to the observer as they occur over time. However, the interpretation of such data must be cautious, since there is a risk of “overinterpreting” the data or making inferences about student behavior from a limited and unstandardized sample of behavior.

The other way of conducting naturalistic observation is the use of A-B-C (*Antecedent-Behavior-Consequence*) observation and recording. The focus here lies in recording the behavior or events occurring just before the behavior of interest is observed (the antecedent) and the behavior or events that are observed as a result of the behavior of interest (the consequence).

An example of this type of observation would be (Hintze et al., 2002): (a) antecedent = a teacher asks some students to take out their paper and pencils; (b) behavior = the target student does not take out their paper and pencil but plays with a toy car on the desk instead; (c) consequence = the teacher reprimands the target student.

Apart from naturalistic observation, there are *systematic direct approaches* to behavioral observation (Hintze et al., 2002). These are characterized by: (a) the goal is to measure specific behaviors, (b) the observed behaviors have been precisely operationally defined, (c) observations are conducted under standardized procedures and are highly objective, (d) the times and places for observation are carefully selected and specified, (e) the scoring and summarizing of the data are standardized and do not vary across multiple observers. The goal of such observation can be, for example, to determine the frequency with which a particular student is out of their seat. First, being out of their seat would be clearly defined¹¹ and then the student would be directly observed for a specified length of time with the number of times he/she got out of his/her seat noted (also, the length of time spent out of their seat might be noted).

Actually, there are several types of data about student behavior, that can be recorded (Hintze et al., 2002): (a) *frequency or event recording* – the observer records the number of occurrences of a behavior observed during a specified time period. It is useful for the behaviors that have a discrete beginning and ending so that their occurrence can be clearly recorded (e.g. raising hands, throwing a pencil, hitting a classmate) and that occur at a relatively low rate; (b) *duration recording* – useful for the behaviors where duration is of importance (e.g. studying, temper tantrums, or social isolation);

¹¹ In Barrish et al. (1969), out-of-seat behavior was operationalized as follows: Leaving the seat and/or seated position during a lesson or scooting the desk without permission. Exceptions to the definition, and instances not recorded, included out-of-seat behavior that occurred when no more than four pupils signed out on the chalkboard to leave for the restroom, when pupils went one at a time to the teacher's desk during an independent study assignment, and when pupils were merely changing their orientation in their seat. Also, when a child left his seat to approach the teacher's desk, but then appeared to notice that someone else was already there or on his way and consequently quickly returned to his seat, the behavior was not counted. Permission was defined throughout the study as raising one's hand, being recognized by the teacher, and receiving consent from her to engage in a behavior. Mitchell et al. (2015) used the following definition: Out-of-seat behavior was defined as the student's buttocks breaking contact with the seat for more than 3 s without a teacher's permission.

(c) *latency recording* – the observer records the elapsed time between the onset of a stimulus or signal (e.g. a teacher’s directive) and the initiation of a specified behavior (i.e. compliance with the directive); (d) *time-sampling interval recording* – with this type of recording, a time period for observation is selected and divided into a number of equal intervals (e.g. a 30 minute observation period can be divided into 180 10-second intervals) and the presence or absence of the target behavior within each interval is recorded.

Table 6

Observations – A Summary of the Strengths and Weaknesses

Strengths	Weaknesses
The examination of behavior and its contextual factors in natural settings	Some behaviors remain hidden to observers
The systematic examination of a behavior’s antecedents and consequences (a good data source for clinicians and psychologists)	High time/personnel cost of lengthy observations
	The risk of misinterpreting the data obtained by naturalistic observations
	The risk of misinterpreting the overall level of school discipline due to short observations
	The risk of the presence of an observer influencing the behavior of the observed individuals

Conducting observations of student behavior definitely plays an important role in both research into school discipline and everyday educational practice. However, several potential limitations can be identified when trying to measure the overall “level” of school discipline (i.e. various types of misbehavior) on a large-scale basis. First, as mentioned by Crothers and Levinson (2004), observation methods may not measure the true prevalence and magnitude of some covert types of misbehavior such as bullying. It frequently occurs in such school areas where there is only a limited opportunity to observe students (e.g. locker-rooms, restrooms). Also, some other types of misbehavior are by their nature “meant” to remain hidden to others such as various forms of academic dishonesty. It could also be argued that the temporal presence of an observer may restrain students from committing some forms of violent behavior towards others. There arises also an ethical issue: what if the observer, who is meant to remain as unobtrusive as possible, becomes a witness to violence between students? His/her interference with the conflict would impact the results of the measurement. However, his/her passive witnessing of the violent conflict where students can be seriously harmed would be, at the very least, disturbing.

Another issue that arises when using observations to measure the level of misbehavior in schools is the scope of observations conducted at a single school. Short-term observations of student behavior in schools might bring biased results, because the number of factors influencing student behavior is enormous (ranging from weather conditions to events occurring in the lesson prior to the observation; Bendl, 2011) and some of them may temporarily change student misbehavior levels, leading researchers to inaccurate conclusions about the overall level of student misbehavior. Long-term systematic observations throughout the school would probably bring more accurate results. However, the time and personnel-consuming nature of such an approach would be immense even at a single school (not to mention for the large-scale measurement of student misbehavior that would allow, for example, an accurate comparison of certain types of schools or schools in different regions). For a summary of the strengths and weaknesses of observations see Table 6.

1.4 Questionnaire surveys (parent-, teacher-, peer-, and self-reports)

Questionnaire surveys are widely used in school discipline research. For example, in research into school bullying, they are the most frequently used data collection method (Wei & Huang, 2005). There are a number of advantages associated with the use of questionnaires. Their low monetary/personnel cost and their easy administration make them particularly preferable for large-scale data collection. Questionnaires are commonly employed to measure the prevalence of various types of student misbehavior in general and to identify those which occur most frequently (e.g. Ding et al., 2008; Koutrouba, 2013), or to measure the prevalence of some specific type of misbehavior such as student academic dishonesty (e.g. Brimble & Stevenson-Clarke, 2005; McCabe, Butterfield, & Trevino, 2006), or bullying (e.g. Kim, Koh, & Leventhal, 2004; Smith & Gross, 2006). What teachers attribute as causes of student misbehavior, the strategies they use to cope with misbehavior, the association between student misbehavior encountered by teachers and the teachers' emotional exhaustion, and many other student misbehavior-related phenomena have been examined using questionnaires (e.g. Ding et al., 2010; Tsouloupas et al., 2010). Questionnaires are also used in intervention studies to determine their effectiveness in reducing student problem behaviors (e.g. Bagley & Pritchard, 1998; Leadbeater, Hoglund, & Woods, 2003). Questionnaires can also be employed to determine

the treatment acceptability of interventions (e.g. Mitchell et al., 2015; Nolan, Filter, & Houlihan, 2014; Wright & McCurdy, 2011).

Questionnaires have also been used frequently in Czech educational research into student school behavior. For example, Bendl used questionnaires to measure the perceived prevalence of various types of misbehavior in schools located in different city areas (2000) and to determine what characteristics students considered a teacher should have to support good classroom discipline (2002). Tomášek (2008) used a questionnaire to measure the prevalence of violence directed at teachers in schools. Vrbová and Stuchlíková (2012) used a questionnaire to measure the prevalence of various forms of dishonest student behavior in schools.

In research into student behavior, we can administer questionnaires to different types of respondents and thus obtain information about the behavior of a particular student from different sources (informants): the student's parents, his/her teacher, his/her peers, or the student him/herself. Multi-informant studies have been conducted in many studies related to student (child) behavior (e.g. Epkins & Meyers, 1994; Fox & Boulton, 2005; McMahon & Washburn, 2003). However, there is often little or only a medium amount of correspondence between the results from the different informants (Achenbach, McConaughy, & Howell, 1987; Branson & Cornell, 2009). In research into student problem behavior, surveys often produce contradictory answers from different informants – students, teachers, principals, parents, trained observers, and the schools' support staff (e.g. Klimusova, Buresova, & Cermak, 2014; Pellegrini & Bartini, 2000).

For example, Pellegrini and Bartini (2000) conducted a comprehensive comparison of various data sources on student aggression and victimization. They showed low to medium levels of correlation between the measurement methods used (observations, teachers' reports, peer-reports, self-reports, diaries). None of the sources correlated more than $r = 0.52$, except for 2 peer-reports. In general, the methods differ in (a) the opportunities and the setting in which the assessor can observe the subject (teacher, peer, parent, trained observer); (b) the relationship to the assessed subject (parent, trained observer); (c) the indicators of the measured trait (overt behavioral clues assessed by an external observer or the respondent's own perspective; adapted from Weiss, Harris, & Catron, 2002). With respect to school discipline research and the use of questionnaires, it is important to

realize that different informants witness students' (school) behavior from different perspectives and may offer different portrayals of a particular student's behavior.

Parent reports

Parents have been rated by mental health professionals as a useful source of information on certain domains of child problem behavior, for example pre-pubertal children internalizing problems (e.g. excessive crying or loss of weight) or conduct problems (e.g. cruelty to animals or running away overnight; Loeber, Green, & Lahey, 1990). However, it must be noted that adults in general are at a distinct disadvantage in terms of the observability of children's behavior, for example in terms of concealed conduct problems (e.g. theft, underage drinking; Loeber et al., 1990). Also, children's problem behavior can differ across settings and can occur exclusively at a school or at home (Loeber et al., 1990). Since parents are usually not present in school they can hardly base an evaluation of many types of their child's school (mis)behavior on their own experience. These limitations make the data on student school misbehavior obtained using parent-reports a less reliable source of information. Also, it might be difficult for the researchers to collect the data since parents are not usually present at a school at one single time (unlike students and teachers) so mass administration would be difficult.

Teacher reports

Teachers appear to be a reliable and practical source of information about student school behavior. In the context of student bullying, Crothers and Levinson (2004) point out that the advantages of teachers' reports are that they are easy to obtain and one teacher can assess a large number of students rapidly. Another advantage is that teachers are often first-hand witnesses (and sometimes even targets) of student misbehavior. A part of their job is to monitor student behavior and be aware of rule violations by students, supporting the notion that teachers are a well-informed data source on student school misbehavior. Indeed, teachers have been rated by mental health professionals as a valuable source of data on child hyperactivity and attention problems (Loeber et al., 1990).

However, it must be noted that teachers usually have limited opportunities to observe particular students. A single teacher can observe a student's

behavior only during his/her own lessons, which make up, especially in the case of older students, only a small part of all the lessons the student attends. The behavior of a student might differ across different subjects taught by different teachers. Hoy and Weinstein (2006) stress that students are not passive recipients of teacher actions, but they choose to resist or comply with rules, ignore, avoid, sabotage, or question teachers' requests. Student behaviors are purposive acts based, among other things, on their relationships with teachers. They summarize that students perceive "good" teachers as worthy of respect, cooperation, and participation. Indeed, other researchers have suggested that some inappropriate teacher behavior might influence student behavior in a negative way (Broeckelman-Post et al., 2016; Kearney et al., 1991). Thus, a teacher's assessment may not reflect the "general" behavior of a student but the behavior of the student in particular circumstances (during particular classes with a particular teacher).

Also, the problem with the "hidden" types of misbehavior that the teachers might not be aware of might negatively affect the accuracy of teacher reports. As mentioned in Pellegrini and Bartini (2000), for example, aggressive acts occur at low frequencies relative to other forms of misbehavior and are usually committed in places and at times when there are few adult witnesses. Overall, administrative ease supports the use of teacher-reports in a large-scale measurement of school misbehavior levels. However, the problems with the accuracy of such reports (hidden cases of misbehavior, limited opportunities to observe students) place severe limitations on the use of such data. We could have multiple teachers assess the same student in the hopes that we obtain a better picture of a student's "general" behavior. However, this would be much more demanding in terms of data collection and still other limitations would remain unaddressed.

Self-reports

Students' self-reports have been used very frequently in research into student behavior, being considered the primary instrument for example in bullying research (Baly, Cornell, & Lovegrove, 2014; Branson & Cornell, 2009). They offer the respondent's perspective and valuable information about phenomena which cannot be (or is only seldom) directly observed by external assessors. The students themselves are best aware of the various types of misbehavior they are committing, including the hidden types (e.g. cheating on exams, bullying others). Also, self-report measures do not

require a great deal of time to administer, they necessitate little manpower, and are inexpensive (Crothers & Levinson, 2004).

However, respondents might be reluctant to admit some types of deviant or taboo behavior and present themselves in a more favorable manner even in anonymous surveys. As mentioned by Branson and Cornell (2009), students may be reluctant to admit to aggression against peers because of the social disapproval associated with being labeled a bully. A similar notion is expressed by Davis, Drinan and Gallant (2009) in the context of research into cheating in schools: we only know what students claim to be doing, not what they are actually doing. The authors believe that in the context of school cheating, students are under-reporting rather than over-reporting their behavior. To sum up, student self-reports appear to be an appropriate way of measuring school misbehavior levels at a large-scale because of the ease of their administration. Also, students themselves have the best knowledge of their own behavior, even those that might not have been observed by any other observers (teachers, peers). A severe limitation, however, is the fact that students might be reluctant to admit to committing socially deviant behavior and might try to present themselves in a more preferable way, even in anonymous surveys, decreasing the accuracy of the results.

Peer reports

When using students as a data source on school misbehavior, we can also make students assess the behavior of their peers – peer-reports. Lindstrom, Lease, and Kamphaus (2007) state that peers provide unique information regarding child behavior when compared to adult raters (parents, teachers). They summarize that peers are very familiar with their classmates, interact with them in a greater number of settings, and have access to a wider array of exchanges than adults. This allows them to have “insider” knowledge of behaviors that are usually hidden from adults, such as bullying. Also Weiss et al. (2002) support the importance of peer perspectives on student school behavior, since students spend a great deal of time in each other’s company (more time than they spend with their parents) during a variety of activities and settings. Brimble and Stevenson-Clarke (2005) also support the value of students as informants about their peers’ behavior. In their study into academic dishonesty they state that students appear to be relatively well informed with respect to the prevalence of dishonest practices among their peers.

Wei and Huang (2005) mention several other advantages of the use of peer-reports in bullying research: (a) evaluating others instead of the self reduces the social desirability issue; (b) if we have a score for a particular student as an aggregate from multiple peers, the reliability of it is often higher than from a single source; (c) peers might be the best informants to assess an individual's involvement in incidents of bullying. Of course, there are some limitations related to the use of peer-reports to measure levels of school misbehavior. For example Weiss et al. (2002) mention that peers might be particularly susceptible to reputation effects, i.e. they make ratings based on a child's reputation rather than on the actual behavior of that child. Also, peer-reports are limited to observable phenomena and are unable to measure students' psychological states such as feelings, attitudes, and beliefs (Wei & Huang, 2005).

The data collection of peer-reports might also be more demanding than self-reports. If every student in a class was evaluated by all his/her classmates, the administrative demands placed on both students and researchers would increase immensely. Wei and Huang (2005) point out that children's interactions in school often extend beyond same-class or same-gender peers. However, obtaining information from their whole school network would be very difficult if not impossible. In summary, there seems to be strong support for the use of students' peers as informants on school misbehavior in large-scale measurements. Also, the social desirability issue occurring with self-reports is reduced and aggregating the score for a particular student based on several peers' assessments might increase the reliability of the score. The limitations of using peer-reports are mainly due to their being restricted to the measurement of observable phenomena, the potential distortion of an assessment by a student's reputation, and the increasing administrative demands when having more peers assess a student. For a summary of strengths and weaknesses of particular informants see Table 7.

Table 7

The Different Strengths and Weaknesses of Teacher, Parent, Self, and Peer-Reports about Student School Behavior

Type of reports	Strengths	Weaknesses
Parent reports	Useful informants on certain types of pre-adolescent behavior problems (e.g. internalizing and conduct problems)	Limited opportunities to observe children's behavior in school (children's behavior might differ across settings) Limited opportunities to observe concealed (hidden) conduct problems More difficult data collection (parents not readily available in schools)
Teacher reports	A single teacher can assess a large number of students rapidly Often a first-hand witness (or a target) of student misbehavior Being aware of student misbehavior is a part of their job	Limited opportunities to observe particular students Assessment does not reflect general student behavior but rather the behavior noticed by the teacher No awareness of hidden types of misbehavior
Student self-reports	Appropriate for large-scale surveys Respondent's unique perspective Information about phenomena which cannot be directly observed by external assessors Good awareness of various types of misbehavior including the hidden types (e.g. cheating on exams, bullying others)	Reluctance to admit to some types of deviant or taboo behavior Presentation of oneself in a more favorable manner
Peer reports	A high degree of familiarity with classmates Access to a wide array of exchanges with classmates Interactions with classmates in a great number of settings Reduction of the social desirability issue (in comparison to self-reports) The ability to assess individual student involvement in rule violations incidents The possibility to aggregate the assessment of a single student based on assessments of several peers	Susceptibility to reputation effects (assessment based on a child's reputation rather than on the actual behavior of the child) Limited to observable phenomena (the inability to measure students' psychological states such as feelings, attitudes, and beliefs) Increasing administrative costs when having more peers assess a single student

Problems with differential scale usage

It must be noted, however, that questionnaires containing items with rating scales share a common limitation, irrespective of which type of respondent is chosen as an informant about student behavior. If we want to use questionnaire surveys to make comparisons in school behavior across individuals, groups of individuals, or countries, it becomes questionable whether respondents' answers are comparable. Results obtained using these questionnaires might be hindered by bias. Bias occurs when the score differences on the indicator of a construct do not correspond to the differences in the underlying trait or ability (van de Vijver & Tanzer, 2004). It may, for example, be the case that two students with the same level of school behavior evaluate their behavior differently – one as excellent, the other, only as good (Vonkova, Bendl, & Papajoanu, 2017). Evidence for differential use of scale has been a long-term concern, not only in education research (Buckley, 2009; Chen, Lee, & Stevenson, 1995; Vonkova, Zamarro, & Hitt 2018) but also in other social sciences research (Bago d'Uva, O'Donnell, & van Doorslaer, 2008; Bago d'Uva et al., 2008; Bago d'Uva et al., 2011; Kapteyn, Smith, & van Soest, 2007; King et al., 2004; Vonkova & Hullegie, 2011). Thus, even though questionnaires offer a relatively cheap and easy way to obtain large-scale data about school discipline, their results must be interpreted with caution (for a summary of strengths and weaknesses of questionnaires see Table 8).

Table 8

Questionnaires – A Summary of the Strengths and Weaknesses

Strengths	Weaknesses
Low monetary/personnel cost and easy administration	The incomparability of results from different respondents due to differential scale usage
Appropriate for large-scale data collection	The limited number of questions in questionnaires (the inability to ask additional questions)

Several techniques have been proposed to adjust for the differential scale usage. One of these techniques, and one which has shown some promising results in educational research, is the anchoring vignette method (AVM). The anchoring vignette method was introduced by King et al. (2004) to adjust self-reports for respondents' heterogeneous reporting style. The basic idea is that respondents first assess themselves. An example of a self-assessment question could be *Overall, how would you assess your school behavior?* with

a five-point scale, where 1 = good behavior and 2, 3, 4, 5 = bad behavior (Vonkova et al., 2017). Secondly, they also evaluate an anchoring vignette(s) – a short story describing hypothetical individuals who manifest the trait of interest. An example of an anchoring vignette related to dishonest student behavior is (Vonkova et al., 2017):

Last month, Honza's class had a substitute teacher two times during their afternoon PE lesson but Honza, on both days went out with his friends instead. He then wrote an absentee note and forged his father's signature. I evaluate Honza's dishonest behavior as a ... (choose a number on the five-point scale).

Since all respondents assess the same anchoring vignette(s), the differences in their answers can be interpreted as differences in scale usage. For example, one student might assess the above-mentioned vignette using the second scale point, while a different student might assess the very same vignette using the third scale point. This information about the heterogeneity in the reporting behavior is then used to adjust self-assessments.

The AVM has been successfully employed in educational research (for a review see Vonkova, Papajoanu, & Bendl, 2016) and has also been employed in the international large-scale survey PISA (student questionnaire) in the years 2012 and 2015. In research into school discipline, the AVM has so far been employed in a single study by Vonkova et al. (2017). The authors studied dishonest student behavior in school. Their findings: (a) demonstrate empirical evidence of heterogeneity in reporting styles across different groups of students when they rate their dishonest behavior; (b) support the further use of the AVM in research into student school behavior.

2 Conclusion

For the large-scale collection of data about student school behavior that would allow the identification of the overall level of school misbehavior in schools, the cost-effectiveness of the method is of the utmost importance to both researchers and practitioners. Questionnaire surveys and school documentation analysis, unlike observations and interviews, allow the collection of data at this scale with reasonable monetary and personnel demands. However, as we have documented above (for more see section 1.1), we believe that school documentation is a very inaccurate data source mainly because of the vague and/or inconsistent standards set for the assessment of

student behavior in schools. If standards for grading students in the domain of school behavior differ across schools, then the comparison of schools based on students' grades could lead us to erroneous conclusions about the misbehavior level at these schools. This is why we believe questionnaire surveys to be the best suited method for large-scale data collection about student school behavior.

As far as questionnaires are concerned, different informants offer different perspectives on student school behavior (for more see section 1.4). We believe that students themselves are a very good source of information about school misbehavior (in comparison to their parents and teachers) because they are not only best aware of their own misbehavior, but also the misbehavior of their peers. We recommend using both student self-reports and student peer-reports in surveys. However, as was previously mentioned (for more see section 1.4), the differences in scale usage among respondents may hinder the results obtained using questionnaire surveys and the comparability of such data.

The combination of student self-reports and peer-reports with the anchoring vignette method seems to be a promising approach to accurately measure student school misbehavior at a large scale. There are also other methodologies that have been proposed to correct for scale usage differences between respondents such as, for example, the identification of the tendencies to select certain scale categories irrespective of the item content (response styles; Baumgartner & Steenkamp, 2001) or the overclaiming technique (Paulhus et al., 2003). It remains a challenge for researchers to investigate the possibilities of using other methods for the correction of differential scale usage in student school behavior research. Of course, it is also possible to use multiple methodological approaches to collect data on student school misbehavior and then triangulate the data obtained using different approaches. However, researchers always must keep in mind the limitations of the various methodological approaches as we have described them in this paper.

References

- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, 101(2), 213–232.

- Bagley, C., & Pritchard, C. (1998). The reduction of problem behaviours and school exclusion in at-risk youth: An experimental study of school social work with cost-benefit analyses. *Child and Family Social Work*, 3(4), 219–226.
- Bago d'Uva T., O'Donnell, O., & van Doorslaer, E. (2008). Differential health reporting by education level and its impact on the measurement of health inequalities among older europeans. *International Journal of Epidemiology*, 37(6), 1375–1383.
- Bago d'Uva, T., van Doorslaer, E., Lindeboom, M., & O'Donnell, O. (2008). Does reporting heterogeneity bias the measurement of health disparities? *Health Economics*, 17(3), 351–375.
- Bago d'Uva T., Lindeboom, M., O'Donnell, O., & van Doorslaer, E. (2011). Education-related inequity in health care with heterogeneous reporting of health. *Journal of the Royal Statistical Association Series A*, 174(3), 639–664.
- Baly, M. W., Cornell, D. G., & Lovegrove, P. (2014). A longitudinal investigation of self- and peer reports of bullying victimization across middle school. *Psychology in the Schools*, 51(3), 217–240.
- Barrish, H. H., Saunders, M., & Wolf, M. M. (1969). Good behavior game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis*, 2(2), 119–124.
- Baumgartner, H., & Steenkamp, J. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38(2), 143–156.
- Bendl, S. (1987). *Prověřování, hodnocení a klasifikace v českém jazyce na 1. stupni základní školy* (Unpublished master's thesis). Prague: Charles university in Prague.
- Bendl, S. (2000). Kázeň na sídlištních školách. *Pedagogická orientace*, 10(3), 54–64.
- Bendl, S. (2002). Které charakteristiky učitele podporují kázeň žáků. *Pedagogika*, 52(3), 346–363.
- Bendl, S. (2011). *Kázeňské problémy ve škole*. Praha: Triton.
- Bohanon, H., Fenning, P., Carney, K. L., Minnis-Kim, M. J., Anderson-Harriss, S., Moroz, K. B., ... Pigott, T. D. (2006). Schoolwide application of positive behavior support in an urban high school. *Journal of Positive Behavior Interventions*, 8(3), 131–145.
- Branson, C. E., & Cornell, D. G. (2009). A comparison of self and peer reports in the assessment of middle school bullying. *Journal of Applied School Psychology*, 25(1), 5–27.
- Brimble, M., & Stevenson-Clarke, P. (2005). Perceptions of the prevalence and seriousness of academic dishonesty in Australian universities. *The Australian Educational Researcher*, 32(3), 19–44.
- Broeckelman-Post, M. A., Tacconelli, A., Guzmán, J., Rios, M., Calero, B., & Latif, F. (2016). Teacher misbehavior and its effects on student interest and engagement. *Communication Education*, 65(2), 204–212.
- Buckley, J. (2009). *Cross-national response styles in international educational assessments: Evidence from PISA 2006*. New York University: Department of Humanities and Social Sciences in the Professions. Retrieved from https://edsurveys.rti.org/PISA/documents/Buckley_PISAresponsestyle.pdf
- Campbell, A., & Anderson, C. M. (2011). Check-in/check-out: A systematic evaluation and component analysis. *Journal of applied behavior analysis*, 44(2), 315–326.

- Casey-Cannon, S., Hayward, C., & Gowen, K. (2001). Middle-school girls' reports of peer victimization: Concerns, consequences, and implications. *Professional School Counseling, 5*(2), 138–148.
- Chen, Ch., Lee, S., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science, 6*(3), 170–175.
- Crothers, L. M., & Levinson, E. M. (2004). Assessment of bullying: A review of methods and instruments. *Journal of Counseling & Development, 82*(4), 496–503.
- Czech School Inspectorate. (2016). *Výroční zpráva České školní inspekce za školní rok 2015/2016*. Retrieved from [http://www.csicr.cz/getattachment/cz/Dokumenty/Vyrocnizpravy/Vyrocnizprava-Ceske-skolni-inspekce-za-skolni-\(2\)/Vyrocnizprava_CSI_2015-2016.pdf](http://www.csicr.cz/getattachment/cz/Dokumenty/Vyrocnizpravy/Vyrocnizprava-Ceske-skolni-inspekce-za-skolni-(2)/Vyrocnizprava_CSI_2015-2016.pdf)
- Czech School Inspectorate. (2017). *Kvalita a efektivita vzdělávání a vzdělávací soustavy ve školním roce 2016/2017: Výroční zpráva České školní inspekce*. Retrieved from http://www.csicr.cz/getattachment/cz/Dokumenty/Vyrocnizpravy/Kvalita-a-efektivita-vzdelavani-a-vzdelavaci-soust/VZ_CSI_2017_web_new.pdf
- Davis, S. F., Drinan, P. F., & Gallant, T. B. (2009). *Cheating in school: What we know and what we can do*. John Wiley & Sons.
- Ding, M., Li, Y., Li, X., & Kulm, G. (2008). Chinese teachers' perceptions of students' classroom misbehaviour. *Educational Psychology, 28*(3), 305–324.
- Ding, M., Li, Y., Li, X., & Kulm, G. (2010). Chinese teachers' attributions and coping strategies for student classroom misbehaviour. *Asia Pacific Journal of Education, 30*(3), 321–337.
- Epkins, C. C., & Meyers, A. W. (1994). Assessment of childhood depression, anxiety, and aggression: Convergent and discriminant validity of self-, parent-, teacher-, and peer-report measures. *Journal of Personality Assessment, 62*(2), 364–381.
- Flower, A., McKenna, J., Muething, C. S., Pedrotty Bryant, D., & Bryant, B. R. (2014). Effects of the good behavior game on classwide off-task behavior in a high school basic algebra resource classroom. *Behavior Modification, 38*(1), 45–68.
- Fox, C. L., & Boulton, M. J. (2005). The social skills problems of victims of bullying: Self, peer and teacher perceptions. *British Journal of Educational Psychology, 75*(2), 313–328.
- Hintze, J. M., Volpe, R. J., & Shapiro, E. S. (2002). Best practices in systematic direct observation of student behavior. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 993–1006). Bethesda, MD: National Association of School Psychologists.
- Hoy, A. W., & Weinstein, C. S. (2006). Student and teacher perspectives on classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice and contemporary issues* (pp. 181–222). Mahwah, NJ: Lawrence Erlbaum.
- iDnes.cz. (2018). *Zastaralé důtky děti ignorují, trestejme i rodiče, navrhuji někteří učitelé*. Retrieved from https://zpravy.idnes.cz/reditelska-dutka-kazenske-opatreni-zakladni-skola-fw-m-/domaci.aspx?c=A180504_161959_domaci_nub
- Institute for Information on Education. (2007). *Rychlá šetření 1/2007*. Copy in possession of authors.
- Institute for Information on Education. (2008). *Rychlá šetření 2/2008*. Copy in possession of authors.
- Irvin, L. K., Horner, R. H., Ingram, K., Todd, A. W., Sugai, G., Sampson, N. K., & Boland, J. B. (2006). Using office discipline referral data for decision making about student behavior

- in elementary and middle schools: An empirical evaluation of validity. *Journal of Positive Behavior Interventions*, 8(1), 10–23.
- Kapteyn, A., Smith, J. P., & van Soest, A. (2007). Vignettes and self-reports of work disability in the US and the Netherlands. *American Economic Review*, 97(1), 461–473.
- Kearney, P., Plax, T. G., Hays, E. R., & Ivey, M. J. (1991). College teacher misbehaviors: What students don't like about what teachers say and do. *Communication quarterly*, 39(4), 309–324.
- Kim, Y. S., Koh, Y. J., & Leventhal, B. L. (2004). Prevalence of school bullying in Korean middle school students. *Archives of pediatrics & adolescent medicine*, 158(8), 737–741.
- King, G., Murray, C., Salomon, J., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 98(1), 567–583.
- Klimusova, H., Buresova, I., & Cermak, I. (2014). Cross-informant agreement and teacher nomination technique in the assessment of children behavior problems. In C. Pracana (Ed.), *Psychology Applications & Developments: Advances in psychology and psychological trends series* (pp. 174–182). Lisboa: inScience Press.
- Koutrouba, K. (2013). Student misbehaviour in secondary education: Greek teachers' views and attitudes. *Educational Review*, 65(1), 1–19.
- Lannie, A. L., & McCurdy, B. L. (2007). Preventing disruptive behavior in the urban classroom: Effects of the Good Behavior Game on student and teacher behavior. *Education and Treatment of Children*, 30(1), 85–98.
- Leadbeater, B., Hoglund, W., & Woods, T. (2003). Changing contexts? The effects of a primary prevention program on classroom levels of peer relational and physical victimization. *Journal of Community Psychology*, 31(4), 397–418.
- Lindsey, B. C. (2008). Looking at positive behavior interventions and supports through the lens of innovations diffusion. *The Innovation Journal: The Public Sector Innovation Journal*, 13(2), 1–18.
- Lindstrom, W. A., Lease, A. M., & Kamphaus, R. W. (2007). Peer- and self-rated correlates of a teacher-rated typology of child adjustment. *Psychology in Schools*, 44(6), 1–21.
- Loeber, R., Green, S. M., & Lahey, B. B. (1990). Mental health professionals' perception of the utility of children, mothers, and teachers as informants on childhood psychopathology. *Journal of Clinical Child Psychology*, 19(2), 136–143.
- Luiselli, J. K., Putnam, R. F., Handler, M. W., & Feinberg, A. B. (2005). Whole-school positive behavior support: Effects on student discipline problems and academic performance. *Educational Psychology*, 25(2–3), 183–198.
- McCabe, D. L., Butterfield, K. D., & Trevino, L. K. (2006). Academic dishonesty in graduate business programs: Prevalence, causes, and proposed action. *Academy of Management Learning & Education*, 5(3), 294–305.
- McCurdy, B. L., Mannella, M. C., & Eldridge, N. (2003). Positive behavior support in urban schools: Can we prevent the escalation of antisocial behavior? *Journal of Positive Behavior Interventions*, 5(3), 158–170.
- McMahon, S. D., & Washburn, J. J. (2003). Violence prevention: An evaluation of program effects with urban African-American students. *The Journal of Primary Prevention*, 24(1), 43–62.

- Mitchell, R. R., Tingstrom, D. H., Dufrene, B. A., Ford, W. B., & Sterling, H. E. (2015). The effects of the good behavior game with general-education high school students. *School Psychology Review, 44*(2), 191–207.
- Morrison, G. M., & Skiba, R. (2001). Predicting violence from school misbehavior: Promises and perils. *Psychology in the Schools, 38*(2), 173–184.
- Nastasi, B. K., & Schensul, S. L. (2005). Contributions of qualitative research to the validity of intervention research. *Journal of School Psychology, 43*(3), 177–195.
- Nolan, J. D., Filter, K. J., & Houlihan, D. (2014). Preliminary report: An application of the Good Behavior Game in the developing nation of Belize. *School Psychology International, 35*(4), 421–428.
- Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology, 84*(4), 890–904.
- Pellegrini, A. D., & Bartini, M. (2000). An empirical comparison of methods of sampling aggression and victimization in school settings. *Journal of Educational Psychology, 92*(2), 360–366.
- Průcha, J. (2005). *Česko-anglický pedagogický slovník*. Praha: ARSCI.
- Smith, R. G., & Gross, A. M. (2006). Bullying: Prevalence and the effect of age and gender. *Child & Family Behavior Therapy, 28*(4), 13–37.
- Sugai, G., Sprague, J. R., Horner, R. H., & Walker, H. M. (2000). Preventing school violence: The use of office discipline referrals to assess and monitor school-wide discipline interventions. *Journal of Emotional and Behavioral Disorders, 8*(2), 94–101.
- Todd, A. W., & Horner, R. H. (2006). *Referral form examples*. Retrieved from http://www.pbis.org/common/cms/files/NewTeam/Data/ReferralFormExamples_v4.0%5B1%5D.doc
- Tomášek, J. (2008). Učitel jako oběť násilí – poznatky z viktimologického šetření na českých středních školách. *Pedagogika, 58*(4), 379–391.
- Tsouloupas, C. N., Carson, R. L., Matthews, R., Grawitch, M. J., & Barber, L. K. (2010). Exploring the association between teachers' perceived student misbehaviour and emotional exhaustion: The importance of teacher efficacy beliefs and emotion regulation. *Educational Psychology, 30*(2), 173–189.
- Van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *Revue Européenne de Psychologie Appliquée, 54*(2), 119–135.
- Vonkova, H., & Hullegerie, P. (2011). Is the anchoring vignettes method sensitive to the domain and choice of the vignette? *Journal of the Royal Statistical Society: Series A, 174*(3), 597–620.
- Vonkova, H., Bendl, S., & Papajoanu, O. (2017). How students report dishonest behavior in school: Self-assessment and anchoring vignettes. *The Journal of Experimental Education, 85*(1), 36–53.
- Vonkova, H., Papajoanu, O., & Bendl, S. (2016). Aplikace metody ukotvujících vinět v pedagogickém výzkumu: přehled literatury a metodologická doporučení. *Pedagogická Orientace, 26*(3), 537–559.
- Vonkova, H., Zamarro, G., & Hitt, C. (2018). Cross-country heterogeneity in students' reporting behavior: The use of the anchoring vignette method. *Journal of Educational Measurement, 55*(1), 3–31.
- Vrbová, J., & Stuchlíková, I. (2012). Školní podvádění starších žáků – pilotní studie. *Pedagogika, 62*(3), 317–331.

- Wei, H. S., & Huang, C. K. (2005). Reviewing school bullying research: Empirical findings and methodical considerations. *NTTU Educational Research Journal*, 16(1), 69–112.
- Weiss, B., Harris, W., & Catron, T. (2002). Development and initial validization of the peer-report measure of internalizing and externalizing behavior. *Journal of Abnormal Child Psychology*, 30(3), 285–294.
- Wright, R. A., & McCurdy, B. L. (2011). Classwide positive behavior support and group contingencies: Examining a positive variation of the Good Behavior Game. *Journal of Positive Behavior Interventions*, 14(3), 173–180.
- Základní škola a mateřská škola Bílá / ZŠ a MŠ Bílá. (2017). *Školní řád*. Retrieved from http://www.zsbila.cz/images/stories/o_skole/2017-2018_Skolni_rad_od_1_9_2017.pdf
- Základní škola a mateřská škola Ostrava – Bělský Les / ZŠ a MŠ Ostrava – Bělský Les. (2017). *Vnitřní klasifikační řád – metodický pokyn*. Retrieved from http://www.zsdvorskeho.eu/index.php?option=com_content&view=article&id=208&Itemid=152
- Základní škola Jana Wericha / ZŠ Jana Wericha. (2011). *Příloha č. 1 Školního řádu ZŠ Jana Wericha – Pravidla pro hodnocení výsledků vzdělávání žáků*. Retrieved from <http://www.zs-jana-wericha.cz/res/archive/067/007335.pdf?seek=1330878920>
- Základní škola Jana Wericha / ZŠ Jana Wericha. (2017). *Příloha č. 2 ke Školnímu řádu ZŠ Jana Wericha – Bodový systém*. Retrieved from <http://www.zs-jana-wericha.cz/res/archive/182/019499.pdf?seek=1506052796>
- Základní škola Petřiny – sever / ZŠ Petřiny – sever. (2015). *Školní řád*. Retrieved from <http://www.zspetriny.cz/soubor-sr-230-.pdf>
- Základní škola Sázavská / ZŠ Sázavská. (2013). *Klasifikační řád*. Retrieved from http://www.zssazavska.cz/data/dokumenty/598_klasifikacni_rad_novy.doc
- Základní škola Želenice / ZŠ Želenice. (2012). *Školní řád*. Retrieved from <https://www.zszelenice.cz/skola12/skolnirad14.pdf>

Authors

Prof. PaedDr. Stanislav Bendl, Ph.D., Charles University, Faculty of Education, Department of Education, Magdalény Rettigové 4, 110 00 Prague 1, Czech Republic, e-mail: stanislav.bendl@pedf.cuni.cz

Doc. RNDr. PhDr. Hana Vonkova, Ph.D. et. Ph.D., Charles University, Faculty of Education, Department of Education, Magdalény Rettigové 4, 110 00 Prague 1, Czech Republic, e-mail: h.vonkova@gmail.com

Mgr. Ondrej Papajoanu, Charles University, Faculty of Education, Department of Education, Magdalény Rettigové 4, 110 00 Prague 1, Czech Republic, e-mail: ondrej.papajoanu@pedf.cuni.cz

Mgr. Eva Vankatova, Charles University, Faculty of Education, Department of Education, Magdalény Rettigové 4, 110 00 Prague 1, Czech Republic, e-mail: aevab@centrum.cz

Analýza metodologických přístupů k výzkumu chování žáků ve školách: problematika neporovnatelnosti žákovského sebehodnocení

Abstrakt: Problematika měření výstupů školního vzdělávání je velice podstatná jak pro výzkumníky, tak pro pedagogy z praxe. Můžeme rozlišit dva hlavní typy výstupů: výstupy v oblasti studijních výsledků (např. v matematice, informačních a komunikačních technologiích či dějepisu) a výstupy v oblasti chování (školní kázeň). Oba typy výstupů jsou ve škole hodnoceny a známkovány. Pokud bychom však nechali různé učitele hodnotit znalosti a dovednosti stejných žáků, jejich hodnocení by se nezřídka lišilo, což se týká i hodnocení chování žáků. Vystává tak následující otázka: Jak přesně měříme výstupy školního vzdělávání? Tato studie má následující cíle: (a) popsat běžně užívané metodologické přístupy k měření chování žáků ve školách a diskutovat jejich výhody a nevýhody. Konkrétně se studie zaměřuje na analýzu školní dokumentace, rozhovory, pozorování a dotazníková šetření. Sekce zabývající se školní dokumentací obsahuje empirickou analýzu školní dokumentace vybraných českých škol; (b) navrhnout inovativní přístup k měření chování žáků ve školách, který by kombinoval žákovské sebehodnocení a žákovské vzájemné hodnocení s metodou ukotvujících vinět s cílem zlepšení porovnatelnosti získaných dat.

Klíčová slova: školní kázeň, studijní výsledky, sebehodnocení, zkreslení, metoda ukotvujících vinět